The Summer Undergraduate Research Fellowship (SURF) Symposium
2 August 2018
Purdue University, West Lafayette, Indiana, USA

Sort vs. Hash Join on Knights Landing Architecture

Victor Pan and Dr. Felix Lin
School of Electrical and Computer Engineering, Purdue University

ABSTRACT

With the increasing amount of information stored, there is a need for efficient database algorithms. One of the most important database operations is “join”. This involves combining columns from two tables and grouping common values in the same row in order to minimize redundant data. The two main algorithms used are hash join and sort merge join. Hash join builds a hash table to allow for faster searching. Sort merge join first sorts the two tables to make it more efficient when comparing values. There has been a lot of debate over which approach is superior. At first, hash join was mainly considered to be faster in most cases, but with the advancements in modern hardware, there is a lot more debate. We look at sort merge vs. hash join on Intel’s Xeon Phi 7210 processor with Knights Landing Architecture. Both algorithms are optimized to utilize the increased hardware capabilities. Our study compares the speed and efficiency of the two algorithms and provides conclusions and recommendations based on our observations.

KEYWORDS

Hash join, Sort merge join, Knights Landing, Intel Xeon Phi, Big Data, Database, Join